Quantitative Trading Series

Quantitative and Qualitative Treatments to Capital Markets

and related bodies of knowledge

By

HangukQuant

Private Notes,

Quantitative Research

2022∼

Quantitative Trading Series

Quantitative and Qualitative Treatments to Capital Markets

and related bodies of knowledge

By

HangukQuant

Private Notes,

Quantitative Research

2022∼

Abstract

This book is designed to be a practical handbook for all finance professionals, practitioner or academic. It is an organization of the various knowledge domains, with a focus on drawing links in the intricate web between the theory and practice of finance that market participants seek to unfold. It contains discussions of trading anomalies, premias and inefficiencies. It contains discussions in discretionary and quantitative trading. Discussion stretches across theoretical work, such as statistical methods, linear algebra and financial mathematics. Applied work in equity research, quantitative research and macroeconomic theory is involved.

This work is attributed to the brilliant writers, academics, scientists and traders before me. Although we have tried to credit the referenced work where relevant, to give a complete reference for its source is impossible. The work has been organized and compiled from various texts, lecture notes, journals, blogs, personal communications and even scraps of scribbled notes from the author's time in college. These contain notes from blogs referencing journals, journals referencing blogs, blogs referencing blogs referring journals - you name it. We apologise if we have failed to credit your work. Other work is original. Readers may reach us at hangukquant@gmail.com. The updated notes are released at `hangukquant.substack.com`.

Faith is to have believe without seeing. This work is dedicated to those who placed their faith in me. To Jeong(s), Choi, Julian and my dearest friends who have shaped my world view and colored it rainbow.

Keywords:

Linear Algebra
Calculus Methods
Computer Methods
Global Macro Trading
Quantitative Research
Statistics & Probability Theory
Risk Premia and Market Inefficiencies
Equities Trading and Other Asset Classes

Table of Contents

Chapter 1

Introduction

1.1 Guidelines for Reviewing Work

The following are the stages of alpha formulations.

Idea 1 (This means to further explore the idea creatively. This is a precursor to a **Test**.).

Test 1 (This refers to parameterized research idea that is to be verified as a **Strategy**.).

Strategy 1 (This explores the implementation and characteristics of a **Test**.).

The following are the stages of theoretical formulations.

Definition 1 (Standard conventions and formal nomenclature are introduced.).

Problem 1 (A formalization of the problem statement is provided).

Exercise 1 (An example or working problem to demonstrate concepts discussed).

The following are stages of theoretical derivations

Lemma 1 (An important result used as is or for other derivations.).

Corollary 1 (An important aside of the theoretical work.).

Theorem 1 (A central result with derivations).

Result 1 (A central result without proof.).

The following are for declarative statements.

Proposition 1 (An opinion of sorts.).

Fact 1 (A statement of (almost) undeniable truth.).

Chapter 2

Ordinary Calculus

Theorem 2 (Integration By Parts)**.** *The integration by parts formula takes form*

$$\int udv = uv - \int vdu \tag{1}$$

Theorem 3 (L'Hopital's Rule)**.**

Chapter 3

Linear Algebra

Here we discuss concepts in linear algebra - notably the literature on this subject is divided into two different schools. One introduces linear algebra as the mathematics and computation of multiply defined linear equations. Here the focus is on teaching linear algebra as a tool for manipulation and computation in multi-dimensional spaces. Determinants are introduced early on, and focuses are on matrix operations. The second approach is to treat matrices as abstract objects, laying focus to the structure of linear operators and vector spaces. Determinants and matrices are only introduced later. Here we provide both - the first will focus on the linear algebraic manipulation of matrices on finite-dimensional, Euclidean spaces. The second treatment will focus on the underlying mathematics of the structure of linear operators and their properties, including the mathematics in infinite dimensional vector spaces and over complex fields.

3.1 Matrix Differentiation

Definition 2 (Differentiation of a Matrix)**.** *Let* $\mathbf{y} = \Psi(\mathbf{x})$ *where* y *is m-vector and* x *is n-vector, then define* $\frac{\delta y}{\delta x}$ *to be the Jacobian matrix of the transformation* Φ*, which is an* $m \cdot n$ *matrix of the first order partial derivatives with elements* ${\frac{\delta y}{\delta x}}_{i,j} = \frac{\delta y_i}{\delta x_j}$ *with* $i \in [m], j \in [n]$*.*

Lemma 2 (Differentiating $\mathbf{y} = \mathbf{A}\mathbf{x}$)**.** *Let* y *be* $m \cdot l$ *matrix and* A *be* $m \cdot n$ *matrix, and* $\boldsymbol{A}$ *is not a function on* $\boldsymbol{x}$*. Then* $\frac{\delta \boldsymbol{y}}{\delta \boldsymbol{x}} = \boldsymbol{A}$*.*

Proof. Since the $y_i = \sum_{k=1}^{n} \alpha_{ik} x_k$ it follows $\frac{\delta y_i}{\delta x_j} = \alpha_{ij}$ for all $i \in [m], j \in [n]$. Then $\frac{\delta \mathbf{y}}{\delta \mathbf{x}} = \mathbf{A}$ □

Theorem 4 (Chain Rule on Matrices)**.** *Let* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$*, and* x *be a function of* $\boldsymbol{z}$*, with* $\boldsymbol{A} \perp \boldsymbol{z}$*. The matrix dimensions are assumed as fit, then*

$$\frac{\delta \boldsymbol{y}}{\delta \boldsymbol{z}} = \boldsymbol{A}\frac{\delta \boldsymbol{x}}{\delta \boldsymbol{z}}.$$

Proof.

$$\frac{\delta \mathbf{y}}{\delta \mathbf{z}} = \frac{\delta \mathbf{y}}{\delta \mathbf{x}}\frac{\delta \mathbf{x}}{\delta \mathbf{z}} = \mathbf{A}\frac{\delta \mathbf{x}}{\delta \mathbf{z}}$$

□

Corollary 2 (General Derivation of the Matrix Differentiation)**.** *By using the results* $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ *and* $(\mathbf{AB})^{\mathbf{T}} = \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$ *together with outcomes obtained in Lemma 2 and Theorem 4 we can derive the differentiation of matrices on more complex linear algebraic equations.*

Chapter 4

Set Theory

4.1 Algebra of Sets

Set theory has many uses under both theoretical and applied settings. One of the topics using sets is probability theory, where probability measure (see Definition 9) and random variables (see Definition 21) are set functions from abstract spaces to real numbers. The most elementary use of sets in probability theory is the treatment of experiments, sample outcomes, sample spaces and events. (see Definitions 4, 5, 6)

Definition 3 (Definition, Operations and Terminologies of Sets)**.** *Here we define common operations that may be applied on sets, collectively known as set algebra.*

1. *Intersection:* $C = A \cap B \implies \forall \epsilon \in C, (\epsilon \in A) \wedge (\epsilon \in B)$.

2. *Union:* $C = A \cup B \implies \forall \epsilon \in C, (\epsilon \in A) \vee (\epsilon \in B)$.

3. *Mutually exclusive (mutex):* $A \cap B = \emptyset = \{\}$.

4. *For* A *on sample space (see Definition 5)* S*, the complement* A^C *is defined as the set satisfying* $(A \cup A^C = S) \wedge (A \cap A^C = \emptyset)$.

A common technique of visualising the relationships between sets is with the use of *Venn diagrams*. In keeping with the dense presentation, we will not present Venn diagrams here. The mistake that many beginners make in relation to sets and probability theory is in using the Venn diagram to infer independence. Do not make the same mistake. Mutual exclusiveness and independence (see Definition 55) are different concepts. Independence involves probability measures (see Definition 9) while mutual exclusion do not require the discussion of probability.

There are two methods in proving that some set $A = B$. The informal method is to draw a Venn diagram and show they represent the same area. The more formal, and mathematically rigorous method is to show that $(A \subset B) \wedge (B \subset A)$. This is done by arguing that for any element $a, a \in A \implies a \in B$ and vice-versa.

Chapter 5

Probability and Statistical Models

The treatment of probability theory is typically done in two-parts, one without the use of measure theory (at the undergraduate level) and the other employing measure theoretic arguments (at the graduate level). Although they are almost never taught together (perhaps for good reason), here we aim to present them shoulder-to-shoulder. Measure theoretic arguments are necessary to draw convergence arguments and discuss concepts belonging to the uncountably infinite world. However, since they are describing the same concept, there is utility in drawing the bridge between the two probability treatments where relevant. Hopefully what is achieved is a more complete view of statistics and probability theory while minimising the attendant disorganization. It is also hoped that this way of presentation also ease the internalization of measure theory probability concepts. We assume basic proficiency in set theory. In fact, not much is assumed, except that the reader is familiar with algebra involving sets. (see Section 4.1)

5.1 Probability Spaces and Probability Measures

Definition 4 (Experiment)**.** *An experiment in statistics is a procedure that can be repeated an infinite number of times with a well defined set of possible outcomes.*

Definition 5 (Sample Space and Sample Outcomes)**.** *Each of the possible outcomes in an experiment (see Definition 4) is called an outcome $s \in S$ where S is the sample space of all possible outcomes.*

Definition 6 (Event)**.** *A set of possible outcomes (a subset of the sample space) (see Definition 5) is known as an event, which we denote e. Then $e \subset S$.*

Inf. probability spaces are used to model situations in which random experiments have infinitely many possible outcomes. There are 2 general types, such as (i) sampling from $x \in [1]$ and (ii) inf. coin tosses. The sample space is the set of possible outcomes. That is $\omega : \omega \in [1]$, and $\Omega_{\text{inf}} = \{\omega = \omega_1\omega_2...,: \omega_n$ represents n-th coin toss$\}$, the set of inf. sequences of head and tails. These sample spaces are both *infinite* and *uncountably infinite*, meaning we cannot list their elements in sequence. The $\mathbb{P}(\omega) = 0$ for any outcome ω belonging to the set. We cannot determine the probability of event $A \in \Omega$ by summation of set members, and define probability of events directly.

Definition 7 (σ-algebra)**.** *Let $\Omega \neq \emptyset$ be a set, let $\mathcal{F}$ be a collection of subsets of Ω. Then $\mathcal{F}$ is σ-algebra (or σ-field) provided:*

1. $\emptyset \in \mathcal{F}$

2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$

3. *sequence* $A_1, A_2, \dots \in \mathcal{F} \implies \cup_{i=1}^{\infty} A_i \in \mathcal{F}$. *Any sequence of sets belonging to* $\mathcal{F}$ *also has union of sets in* $\mathcal{F}$.

It is easy to confuse the *power set with the sigma algebra of a set. The power set is the largest possible sigma algebra of a set.* The trivial σ-field $\{\emptyset, \Omega\}$ is not the power set of Ω, but is also σ-algebra. All operations on elements of sigma algebra set gives us other sets in the sigma algebra. It is easy to derive that any union of subsets are in the sigma algebra, and so are their intersections (consider the De-Morgan on complement of the union of complements).

Definition 8 ($\mathbb{P}$ function)**.** *The probability function is a function* $\mathbb{P}$ *that satisfies axioms that are collectively known as 'Kolmogorov Axioms of Probability'. For sample space* Ω *and events* $A \in \mathcal{F}$ *defined over the sample space, the* $\mathbb{P}$ *function satisfies the same properties as defined in Definition 9.*

Definition 9 ($\mathbb{P}$ measure)**.** *Let* Ω *be non-empty set and* $\mathcal{F}$ *be* σ*-algebra of subsets of* Ω*. Then probability measure* $\mathbb{P}$ *is function mapping every set* $A \in \mathcal{F}$ *to range* $[0, 1]$*, written* $\mathbb{P}(A) : \mathcal{F} \to [0, 1]$*. We require*

1. $\mathbb{P}(\Omega) = 1$, *note that* $\Omega \in \mathcal{F}$ *given* $\emptyset \in \mathcal{F}$ *and complement property.*

2. *(countable additivity) where* $A_1, A_2, \cdots$ *are disjoint set sequence, then* $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \Sigma_{n=1}^{\infty} \mathbb{P}(A_n)$*. This implies finite additivity* $\mathbb{P}(\cup_{i=1}^{n} A_i) = \Sigma_{i=1}^{n} \mathbb{P}(A_i)$ *on disjoint sets.*

Theorem 5 (Properties of Probability Measures/Functions)**.** *We state some trivial but important results of probability functions without proof. Beginning with (i)* $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$*, (ii)* $\mathbb{P}(\emptyset) = 0$*, (iii)* $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$*, (iv)* $\mathbb{P}(A) \leq 1$*, (v)* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$*. A less trivial theorem is provided with proof, with regards to the probability measure on k unions:*

$$\mathbb{P}(\cup_i^n A_i) = \sum_i^n \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n+1} \mathbb{P}(\cap_i^n A_i) \qquad \textit{for any } A_i \textit{ on } S,\ i \in [n]. \tag{2}$$

Proof. (verify this) □

Exercise 2 (Math is Weird Sometimes)**.** *Consider events* A, B*. Then we have* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - (1 - \mathbb{P}((A \cap B)^C) = \mathbb{P}(A) + \mathbb{P}(B) - (1 - \mathbb{P}(A^C \cup B^C)) = \mathbb{P}(A) + \mathbb{P}(B) - 1 + \mathbb{P}(A^C \cup B^C)$*. The probability of at least one of* A, B *occurring increases with the probability of at least one not occurring!*

Exercise 3. *Consider a probability measure as defined in Definition 9, then prove that*

1. $(A \in \mathcal{F}, B \in \mathcal{F}) \wedge A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

2. $A \in \mathcal{F} \wedge \{A_n\}_{n=1}^{\infty} \in \mathcal{F} \wedge \lim_{n\to\infty} \mathbb{P}(A_n) = 0 \wedge (\forall n, A \subset A_n) \implies \mathbb{P}(A) = 0$

Proof.

1. We see that for $A \subset B$ we have $B = A \cup (B \backslash A)$ the result follows by countable additivity of disjoint sequences.

2. For all n, $\mathbb{P}(A) \leq \mathbb{P}(A_n)$ and therefore $\mathbb{P}(A) \leq \lim_{n\to\infty} \mathbb{P}(A_n) = 0$ and $0 \leq \mathbb{P}(A) \leq 0$.

□

Exercise 4. *Prove that the set of sequences of coin tosses in which outcome of each even numbered coin toss matches the outcome of the preceding toss, such that*

$$A = \{\omega = \omega_1 \omega_2 \cdots : \omega_1 = \omega_2, \omega_3 = \omega_4\}$$

is uncountably infinite. Furthermore, show that if p-head is not zero or one, then $\mathbb{P}(A) = 0$.

Proof. Consider the function $\phi : A \to \Omega$ and $\phi(\omega) = \omega_1\omega_3\omega_5\cdots$ then the function is injective and surjective. Then the cardinality of A matches cardinality of Ω_∞, which means that A is uncountably infinite. Next, let $A_n = \{\omega : \omega_1 = \omega_2, \cdots \omega_{2n-1} = \omega_{2n}\}$. Then

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \left[p^2 + (1-p)^2\right]^n \overset{p\notin\{0,1\}}{=} 0.$$

□

Definition 10 (Probability Space)**.** *The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is termed probability space, with reference to definitions 7 and 9.*

Definition 11 (Uniform Lebesgue Measure $\mathcal{L}$ on Unit Intervals)**.** *Models choice of sampling a random number from unit interval, with probability measure on $[a, b]$ by*

$$\mathbb{P}[a, b] = \mathbb{P}(a, b) = b - a, \qquad 0 \leq a \leq b \leq 1$$

. Probability measure on point is 0*.*

Consider that we can define an open interval (a, b) as a union of sequence of closed intervals, we can write

$$(a, b) = \cup_{n=1}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n}\right],$$

. Now consider a σ-algebra formed by starting with the closed *intervals* and putting in everything else required to have a σ-algebra. Then it turns out that this σ-algebra contains all open intervals also.

Definition 12 (Borel σ-algebra, $\mathcal{B}$)**.** *The σ-algebra constructed by beginning with closed intervals and adding everything else required to have a σ-algebra is called the Borel σ-algebra of subsets $[0, 1]$, denoted $\mathcal{B}$. Sets belonging to the set $\mathcal{B}$ are called Borel sets, and are subsets of $[0, 1]$.*

Exercise 5 (Infinite, Independent Coin Toss Space)**.** *We can illustrate infinite probability spaces with a sequence of infinite coin tosses. Let Ω_∞ denote the set of possible outcomes, and the probability of head, tail be $p, q = (1 - p)$ respectively, both non-zero. The tosses are independent, and we want to construct a probability measure on this space corresponding to this experiment. We can define $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$. The 2 sets form a σ-algebra, and we denote this $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Note that $|\mathcal{F}_0| = 2^{2^0}$. Then, consider the 2 sets A_H and A_T, denoting the sets that begins with a head and the other with a tail. For instance, A_H may be denoted $\{\omega : \omega_1 = H\}$. Let the $\mathbb{P}(A_H) = p, \mathbb{P}(A_T) = q$, and we have defined the probability measure $\mathbb{P}$ for the σ-algebra $\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}$. Note also that $|\mathcal{F}_1| = 2^{2^1}$. No other sets need to be added to form a σ-algebra. We can continue for $\mathcal{F}_2$ of size 2^{2^2} and so on. The continuation of this process gives us the probability of every set that can be described in terms of finitely many tosses. It turns out that once this is done, the other sets that are not describable in terms of finitely many coin tosses have determined probabilities. Consider for example, consider the set containing singleton infinite sequence of heads $HHH\cdots$, which is not a finite sequence but is a subset of $A_H, A_{HH}\cdots$. Furthermore, since we defined $\mathbb{P}(A_H) = p, \mathbb{P}(A_{HH}) = p^2$ and so on, the singleton set has probability equals zero as $p < 1$. The same argument can used to argue that the probability of any individual sequence $\in \Omega_\infty$ equals zero. We create the σ-algebra $\mathcal{F}_\infty$ by putting in every set that can be described by the finite coin tosses, and then everything else required for the σ-algebra property, and it turns out that we will then have the probability of every set in $\mathcal{F}_\infty$. It is determined but not necessarily easily computed, as we shall see. For instance, consider the set $A = \{\omega = \omega_1\omega_2\cdots : \lim_{n\to\infty} \frac{H_n(\omega_1\cdots\omega_n)}{n} = \frac{1}{2}\}$, which defines the set for which the long-run average of heads is half. This is in $\mathcal{F}_\infty$. To see this, for constant $m, n \in \mathbb{Z}^+$ define*

$$S_{n,m} = \{\omega : \left|\frac{H_n(\omega_1, \cdots, \omega_n)}{n} - \frac{1}{2}\right| \leq \frac{1}{m}\}$$

. This is in $\mathcal{F}_n$ with known probability. By definition of limit, the specified limit is satisfied iff *for every positive integer m, there exists a positive integer N such that $\forall n \geq N, \omega \in S_{n,m}$. In other words, the set for which ω satisfies the limit can be expressed*

$$A = \cap_{m=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} S_{n,m}.$$

Then $A \in \mathcal{F}$ by formulation since it is the union and intersection of members of the set. It turns out the Strong LLN asserts $\mathbb{P}(A) = 1$ if $p = 0.5$ and 0 otherwise.

The probability zero event in uncountable probability spaces has a paradox, as highlighted in our example above. Whenever an event is said to be *almost sure*,we refer to it as the case $\mathbb{P}(A) = 1$, even though it may not include every possible outcome. The events not included together has probability $\mathbb{P}(A^c) = 0$.

Definition 13 (Almost Surely)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say event A occurs* almost surely*.*

5.2 Counting and Combinatorics

We have defined probabilities, sample spaces, events and other important fundamental artefacts of randomness in the Section 5.1. The most basic probability model is the counting model. Combinatorics is the science and practice of counting, arranging and ordering of objects. Surprisingly, like Kolmogorov probability axioms, combinatorics can be reduced to four fundamental rules.

Theorem 6 (Combinatorial Multiplication Rule)**.** *Operations A_i, $i \in [k]$ performed in sequence can be conducted in total of $\Pi_i^k n_i$ where n_i is the number of ways to conduct A_i.*

Theorem 7 (n-permutate k)**.** *The number of permutations of length k from n distinct objects without repetitions is written*

$$^{n}P_k = \frac{n!}{(n-k)!}. \tag{3}$$

Proof. The proof follows from application of Theorem 6 by first considering n possible ways, then $n-1$ ways down to $n-k+1$ ways. □

Corollary 3. *By the Theorem 7 the number of ways to arrange n distinct objects is $n!$.*

Result 2 (Stirling's Approximation)**.** *Stirling's formula for approximating $n!$ is written*

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} \exp(-n). \tag{4}$$

In practice we can write in log form

$$\log \sqrt{2\pi} + (n + \frac{1}{2}) \log n - n$$

and then exponentiating after substituting the value of n.

Exercise 6 (The Rook Problem)**.** *How many ways are there to arrange eight rooks on an eight by eight chessboard such that they are non capturing? See that we each rook once placed, eliminates one row and one column each from the next iteration. Eight rooks can have valid formations $^{8}P_8$ and be internally permutated $^{8}P_8$ times for a total of $8! \cdot 8!$ arrangements.*

Definition 14 (n-permutate r categories, multinomial coefficients)**.** *The number of permutations of n objects of r categories where each type i has n_i, $i \in [n]$ objects is written*

$$\frac{n!}{\Pi n_i!} \tag{5}$$

where $\sum_i^r n_i = n$. This is known as the multinomial coefficient, due its appearance when expanding multinomials.

Exercise 7 (Larsen and Marx [4])**.** *Find the coefficient of x^{23} in the expansion of $(1 + x^5 + x^9)^{100}$ by combinatorial arguments.*

Proof. The coefficient of the term corresponds to the number of ways in which the term can be formed. Here x^{23} can only be formed from two of x^9, one of x^5 and ninety seven ones being multiplied together. Then the coefficient is $\frac{100!}{2!1!97!}$. □

Exercise 8 (Number of Passwords)**.** *How many total passwords can be constructed that is length of ten, constructed from four letters, four numbers and two symbols. Let the total symbols admissible be eight.*

Proof. One can choose 10^4 total numbers, 8^2 symbols and $26^4 \cdot 2^4$ letters (including upper cases). Once the numbers, letters and symbols are chosen, they can be arranged in $\frac{10!}{4!4!2!}$ ways. The total number of passwords is then $\frac{10!}{4!4!2!} \cdot 10^4 \cdot 8^2 \cdot 26^4 \cdot 2^4$. □

See that we can often think of the permutation problem as a two step choice of first choosing the candidates and then arranging them.

Theorem 8 (Circular Permutations)**.** *There are $(n-1)!$ ways to permute n distinct objects in a circle. Do this by writing ${}^nP_n$ and see that a factor of n permutation arrangements are repeated by the 'circularity'. Divide by n and we get $n!/n = (n-1)!$.*

Exercise 9 (The Necklace Problem)**.** *We have 10 beads of different colours, how many different necklaces can we form?*

Proof. Their circular arrangement permutation cardinality is $(n-1)!$ for $n = 10$. However, since the necklace flipped over is a different circular permutation but the same necklace, we actually need to divide by two to obtain $\frac{(n-1)!}{2}$. To see this consider a smaller necklace of ROY. This is the same necklace RYO flipped! We divide by n to account for ROY being identical to OYR and YRO. The flipping accounts for further division factor of two. □

Definition 15 (n-choose k, binomial coefficients)**.** *The number of ways to form combinations of size k from set of n distinct objects without repetitions is written*

$$\binom{n}{k} = {}^nC_k = \frac{n!}{k!(n-k)!}. \tag{6}$$

Due to its common appearance as coefficients in binomial expansions, this term is also called the binomial coefficients. Cross reference this with the multinomial coefficients. (see Definition 14)

Proof. The proof for this can be seen by first permuting k of n objects to get ${}^nP_k$ and then seeing that order does not matter and dividing by $k!$. □

Theorem 9 (Pascal's rule). *For positive, natural numbers n and k, the Pascal's rule states that $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$. This can be thought of recursively. To choose k objects from $n+1$ objects, I can first choose or not choose the 'first' object. Choosing the object will mean I can choose $k-1$ objects from the remaining n objects. If I do not choose the object then I still have to pick k out of the remaining n objects. The total number of possibilities are the sum total.*

Exercise 10. *See in Theorem 5 we defined the probabilities of n unions of events. Here we want to show that the formula adds the probability of each outcome exactly once - there is no double counting or under counting. Consider the set of outcomes in $\cup_i^n A_i$ that belong to some k of the A_i and no others. We want to show that the formula counts this set of outcomes exactly once for arbitrary k. See that these outcomes get counted $\binom{k}{1}$ times in term $\sum_i^n \mathbb{P}(A_i)$, $\binom{k}{2}$ times in $\sum_{i<j} \mathbb{P}(A_i \cap A_j)$ and so on. The outcomes are counted a total of*

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} \cdots + (-1)^{k+1}\binom{k}{k} \tag{7}$$

times. See that we can write the binomial expansion $(-1+1)^k = 0^k = 0 = \sum_{j=0}^k \binom{k}{j}(-1)^j 1$. Then see that Equation 7 equals $\binom{k}{0} = 1$ and we are done.

Theorem 10 (Some Binomial Identities). *Prove that*

1. $\sum_{i=1}^n i \cdot \binom{n}{i} = n \cdot 2^{n-1}$.
2. $\sum_i^n \binom{n}{i}^2 = \binom{2n}{n}$.
3. $\sum_{k=0}^n \binom{n}{k} = 2^n$.
4. $\binom{n}{k} = \binom{n}{n-k}$.

Proof. For 1. see that $(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k \cdot 1$. Differentiating both sides we get $n(1+x)^{n-1} = \sum_{k=0}^n \binom{n}{k} k x^{k-1}$. Substitute $x = 1$ and we are done. For 2. logicize that

$$\binom{2n}{n} = \binom{n}{0}\binom{n}{n} + \binom{n}{1}\binom{n}{n-1} + \cdots + \binom{n}{n}\binom{n}{0}.$$

Using the identity $\binom{n}{k} = \binom{n}{n-k}$ we obtain the equality. □

We have worked hard to count the total number of arrangements and choices of objects. This is often motivated by the desire to find the probability of an event.

Theorem 11 (Combinatorial Probability, the classical definition of probability). *If there are n ways to perform an operation and m satisfies some condition for occurrence of an event, then $\mathbb{P}(A) = \frac{m}{n}$.*

Exercise 11 (The Birthday Problem). *Assume that birth is uniformly distributed over 365 days in a year and k people are selected at random. What is the probability that there is at least one overlap in birthday dates?*

Proof. By multiplication rule (see Theorem 6) the total number of birthday sequences are 365^k. The total number of sequences with k people with distinct birth dates are ${}^{365}P_k$. Then the probability that at least two people share birthdays are simply

$$\frac{365^k - {}^{365}P_k}{365^k}. \tag{8}$$

We require $n \geq 23$ for the probability to be greater than half. □

Exercise 12 (Discrete Random Walk). *A drunkard walks forward and backward randomly with equal probability. At time/step n what is the probability that he is r steps ahead of where he began?*

Proof. Let x be number of forward steps and y be number of backward steps. Then $x + y = n$ and $x - y = r$ and the equations solve to $x = \frac{n+r}{2}$, $y = \frac{n-r}{2}$. The total number of ways for which he ends up at r-front is $\frac{n!}{\frac{n+r}{2}!\cdot\frac{n-r}{2}!}$ and the total number of ways to take n steps is 2^n. His probability for r-forward is $\frac{\binom{n}{\frac{n+r}{2}}}{2^n}$. □

Definition 16 (von Mises probability). *For an experiment (see Definition 4) repeated n times under identical conditions and if event E (see Definition 6) occurs m times out of the n repetitions, the probability of event E is written $\mathbb{P}(E) = \lim_{n\to\infty} \frac{m}{n}$.*

5.3 Random Variables

In general, a random variable can be said to be a function that maps outcomes from a sample space to real numbers. We write notations such as $X : \Omega \to \mathbb{R}$ to show this. When the sample space Ω is finite, or countably infinite we can define a probability function as such:

Definition 17 (Probability Function on Countable Sample Space). *For finite or countably infinite sample space Ω, real valued function p is said to be discrete probability function if it satisfies $0 \le p(s), \forall\omega \in \Omega$ and $\sum_{\omega\in\Omega} = 1$.*

Definition 18 (Discrete Random Variable). *A function with domain Ω ranging over finite or countably infinite set of real numbers is called a discrete random variable. Then for such a random variable X we can write $X(\omega) = k$ for $\omega \in \Omega$ and $k \in \mathbb{R}$.*

Definition 19 (Discrete Probability Density Function). *Associated with the random variable defined as in Definition 18, its probability density function $p_X(k)$ is the probability function such that*

$$p_X(k) = \mathbb{P}(\{\omega \in \Omega | X(\omega) = k\}) = \mathbb{P}(X = k). \tag{9}$$

The probability density function for a discrete random variable (Definition 18) can be expressed by formula or by enumerating the domain (in the finite case) and assigning probability values. Often we are interested in the probabilities of a range, as opposed to at a point. We can then consider its cumulative density function.

Definition 20 (Discrete Cumulative Density Function). *For discrete random variable X (Definition 18), the probability that X takes on values t or lesser is written by its cumulative density function F_X, written*

$$F_X(t) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \le t\} = \mathbb{P}(X \le t). \tag{10}$$

Random variables often take values that have continuous domain. We might have sample spaces that contain an uncountably infinite number of outcomes. A discrete probability function $p(s)$ as defined in Definition 17 is not applicable to outcomes in continuous sample spaces. Uncountably infinite sample spaces often have point probabilities of zero. We give a more general definition under measure theoretic settings that generalizes the behavior of random variables.